

A Network Biology Approach to Understand Combination of Drugs

Ke-Jia Xu^{1,2} Jiangning Song³ Xing-Ming Zhao^{2,*}

¹Department of Mathematics, Shanghai University, Shanghai 200444, China

²Institute of Systems Biology, Shanghai University, Shanghai 200444, China

³Department of Biochemistry and Molecular Biology, Monash University, Clayton,
VIC 3800, Australia

Abstract Drug combination is becoming a popular strategy for combating complex traits, where combination of drugs improves therapy significantly over single agents. However, it is a challenging task to identify the optimal and effective combinations of drugs due to explosion of huge number of possible combinations among drugs. Therefore, it is necessary to understand the mechanisms underlying drug combinations. In this work, a network biology approach is presented to investigate the combination drugs and their target proteins in the context of protein interaction network, human pathway and functional annotations, in order to better understand the underlying rules of drug combinations. Our results indicate that the target proteins of effective combination drugs tend to be close in the protein interaction network, are more likely to be involved in the same pathways and their target proteins tend to have similar biological functions. These findings shed lights on the underlying principles of drug combinations and are helpful for identifying possible effective combinations of drugs.

Keywords Drug combination, Gene Ontology, pathway, protein-protein interaction

1 Introduction

Drug discovery is a long term and expensive procedure, which in history generally produces ‘one-target one-disease’ drugs that fail to work to treat complex diseases, e.g. cancer [1, 2, 3]. Recently, drug combination, i.e. the combination of approved drugs, is becoming a popular strategy in clinic, where combination of drugs improves therapy significantly over single agents [4, 5]. For example, Avandamet is the combination of Metformin and Rosiglitazone, which is an approved combination drug to combating Type 2 Diabetes [6]. Agrawal et al. found a combination drug for treating Huntington disease based on experiments in *Drosophila* [7]. However, most of combination drugs in use are found by experience. That is, the combination drug works as a blackbox and the mechanism of drug combination is not known. In a combination, one drug may promote or suppress the effect of another one, and two drugs may produce a different one from any individual drugs. The number of possible drug combinations will increase exponentially with more drugs available. Considering the search space of possible combinations

*Corresponding author, Email: xm_zhao@shu.edu.cn, Tel:86-21-66136130, Fax: 86-21-66136129.

of known drugs, it is a challenging task to identify optimal and effective combinations of drugs. Therefore, it is impossible to screening the effective drug combinations by experiments. Most recently, we have developed a computational framework for predicting drug combinations [8]. However, this method works mainly based on gene expression profiles generated under perturbations induced by drugs, which are generally not publicly available, especially for combination of drugs.

In general, one drug binds to its targets which in turn affect downstream pathways or biological processes. In other words, drugs work by affecting some biological processes that are perturbed by corresponding target molecules. In the case of combination drugs, combination of different drugs that have different action mechanisms targeting different biochemical pathways or interaction networks has more potential to be applied to shut down the multi-factor controlled processes of complex diseases. Therefore, the network circuits in which the drug target proteins work should provide useful insights into the action mechanisms of the combination drugs. In this work, a new network biology approach is presented to investigate the combination of drugs. In particular, the target proteins are investigated in the context of protein interaction map and pathway circuits associated with combination drugs. Our results demonstrate that the target proteins of combination drugs are distributed differentially from those of non-effective combinations, which can help to predict possible drug combinations in the future.

2 Results

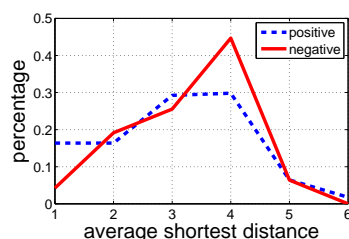


Figure 1: The effect radius of drug combinations.

Firstly, the effect radius was investigated for positive samples and negative samples. Figure 1 shows the ratio of target proteins over certain threshold. It can be seen that positive samples tend to bind to target proteins that are closer neighbors in the protein interaction network, in contrast to the negative samples.

Since the approved drugs can be used as combination drugs, their components were studied to show whether they share more target proteins compared with non-effective combinations. Figure 2 shows the percentage of drug combinations versus the ratio of target proteins that are shared by at least two drug components for positive and negative samples. In fact, only about 20% of drug combinations share at least one target protein. It can be seen that the effective drug combinations tend to share more target proteins than negative combinations.

Furthermore, the interaction partners of shared target proteins were investigated. Here,

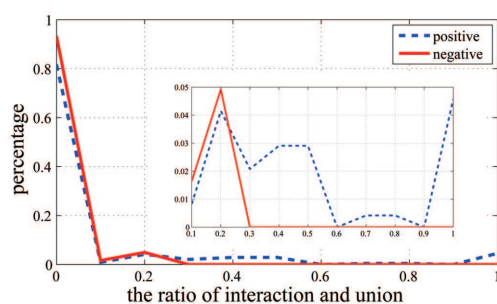


Figure 2: The percentage of drug combinations versus the ratio of target proteins that are shared by at least two drug components. Since only few drug combinations have shared target proteins, the range from 0.1 to 1 is amplified.

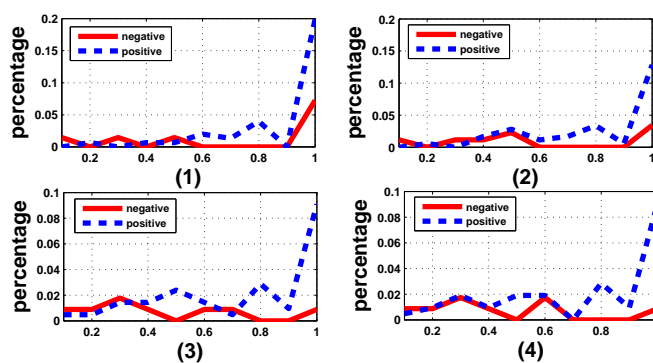


Figure 3: The percentage of drug combinations whose drug components have different interaction neighborhood. The target proteins that can be reached in 2, 3, 4 and 5 steps are regarded as the interaction partners for the four graph, respectively.

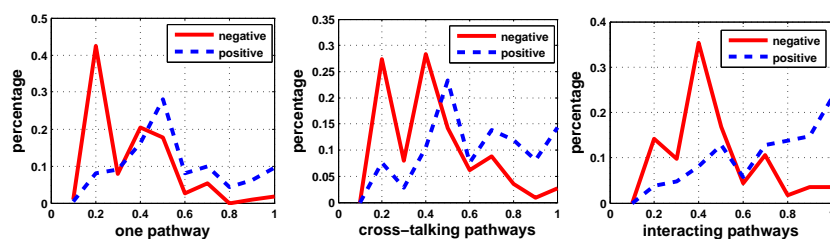


Figure 4: The percentage of drug combinations versus the ratio of target proteins respectively involved in one pathway, cross-talking pathways, and interacting pathways.

the target proteins that can be reached in five steps by shared target proteins are regarded as their interaction partners, where the shortest path can be treated as a pathway. Figure 3 shows the percentage of drug combinations whose drug components have different interaction neighborhood. It can be seen that the target protein tend to be tightly linked for positive combinations compared with negative ones. The results imply that for effective combination of drugs, the target proteins should be close in the protein interaction network, which is reasonable considering the shorter distance between target proteins actually means few side effects induced by the downstream pathway of target proteins.

Figure 4 shows the percentage of drug combinations versus the ratio of target proteins respectively involved in one pathway, cross-talking pathways, and interacting pathways. For the same pathway, the positive peak (at 0.5) is located right to the negative peak (at 0.2), which implies that the targets of positive combinations are more likely in one pathway. The similar phenomena are found for both cross-talking pathways and interacting pathways. Since the more pathways affected by a drug, the more possible side effects may be introduced by the drug. The results demonstrate that positive combination drugs can have few side effects.

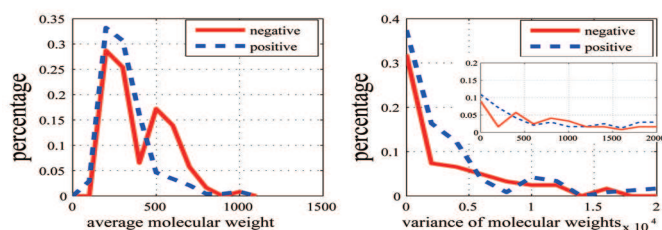


Figure 5: (1)The average molecular weight of drug components from the drug combination. (2)The variance of molecular weights for a drug combination .

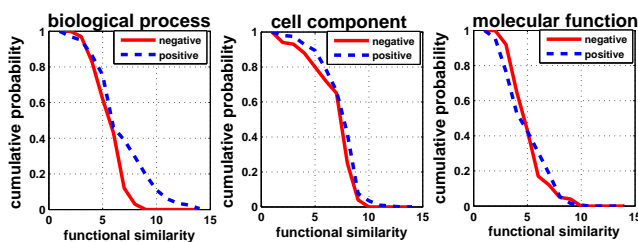


Figure 6: The distribution about GO similarity of drug targets for biological process , cell component, and molecular function, respectively. For example, in the first figure, the element with coordination (8, 0.3) means that there are 30% drug combinations with GO similarity larger than 8.

In addition, the molecular weights were investigated for drug components from a combination. The molecular weight of each drug was obtained from DrugBank, and the average molecular weight was calculated for each drug combination. Figure 5 show the average and variance of molecular weights for drug components from the drug combina-

tion. From the analysis, we can see that the drug components in the positive combinations tend to have small molecular weights and similar molecular weights.

Since the drugs affect the biological system through their target proteins, the function of target proteins can help to understand the actions of drug combinations. Figure 6 shows the percentage of drug combinations versus different GO similarities. From this figure, we can see that positive combinations tend to have target proteins that are involved in similar biological processes, which is consistent with the pathway analysis for drug targets.

In addition, the independent-samples T test is used to compute the p-value of these results. It is worth mentioning that all the distributions displayed in Figure 2-6 are statistically significant (p-values are far less than 10^{-5}), except that the distribution of average distance (Figure 1, with p-value of 0.081).

3 Materials and Methods

3.1 Drug combinations

The drug combinations are retrieved from a newly released Drug Combination Database (DCDB) [9], which collects effective drug combinations from literature. The target protein information is obtained from DrugBank [10]. The drug combinations that do not have target information for corresponding drug components are discarded. Finally, 241 combination drugs are retained and used as positive samples. The drug interaction information for the drugs involved in the 241 drug combinations is obtained from DrugBank, and a negative set is constructed which contains those drug combinations that are annotated to interfere with each other. At last, 122 drug combinations are collected in negative dataset.

3.2 Molecular network and protein annotations

The human protein-protein interactions (PPIs) from BioGRID database [11] are used in this work, which contain 2633 interactions after deleting the self-interactions and the duplicate ones. The human pathway information is retrieved from KEGG database [12].

To investigate the functional information of drug target proteins, the annotations from Gene Ontology (GO) database are used, which covers three domains, including cellular component, molecular function, and biological process [13]. The annotations for human proteins are extracted from GO database for further analysis.

3.3 The effect radius of combination drug

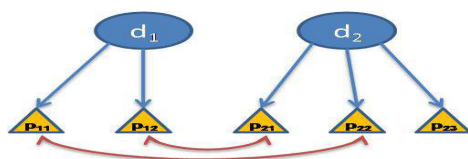


Figure 7: The distance between two drugs d_1 and d_2 . The ellipses represent drugs, and the triangles represent target proteins. The red lines represent the shortest paths between drug targets in PPI.

Given the drug components in a combination drug and their corresponding target proteins, the distance $dis(i, j)$ between two drugs i and j is defined as the average of distances

between target proteins of i and those of j , where the distance between two target proteins is the shortest path distance obtained based on the protein-protein interaction network. For example, in Figure 7, the red lines represent the shortest paths between targets of drug d_1 and those of drug d_2 , and $dis(1, 2)$ is the average of all distances accompanying red lines. With the distance between any two drugs available, the effect radius of a combination drug can be defined as below.

$$R = \frac{\sum_{(i,j) \in D, i \neq j} dis(i,j)}{|D|} \quad (1)$$

where R is the effect radius of one combination drug, $dis(i, j)$ denotes the distance between drugs i and j , D is the set of all combinations among components of a drug combination, and $|D|$ is the size of D .

3.4 The interaction neighborhood of combination drug

The drug components in a combination drug sometimes share target proteins. To see how such kind of combination drugs work, a similarity score $s(i, j)$ is defined for two drugs as $s(d_i, d_j) = \frac{I(d_i, d_j)}{U(d_i, d_j)}$, where $I(d_i, d_j)$ and $U(d_i, d_j)$ represent the number of intersection and union of targets of drugs i and j , respectively. For combinations with n ($n > 2$) drug components, the similarity among drugs can be described as follows.

$$S = \max_{1 \leq i, j \leq n} (s(d_i, d_j)) \quad (2)$$

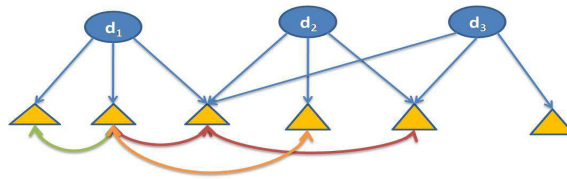


Figure 8: The red edges are links between common targets. The orange edges are ones linked with proteins from different drugs.

In addition, the interaction partners of targets shared by at least two drugs are investigated. Suppose one drug combination has n drug components, the number of edges between all drug targets is m (all edges in Figure 8), the number of edges for common target proteins is k (red edges in Figure 8), and the number of shared proteins by different drugs is s , we define an interaction neighborhood $C = \frac{k+s}{m+s}$ for each combination drug.

3.5 The interaction between pathways and combination drug

In general, the target proteins from two drugs can be grouped into one pathway, cross-talking pathways, interacting pathways as shown in Figure 9 if these proteins have functional relationships. The interaction between a combination drug and a pathway is defined as the ratio of target proteins that are involved in the pathway. Finally, the pathway that has the maximum ratio of target proteins is regarded as the pathway affected by the drug, and this ratio is used to investigate the difference between positive and negative samples.

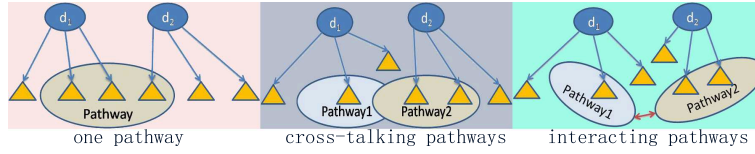


Figure 9: The distribution of drug target proteins in context of pathways.

3.6 Functional enrichment of target proteins of combination drug

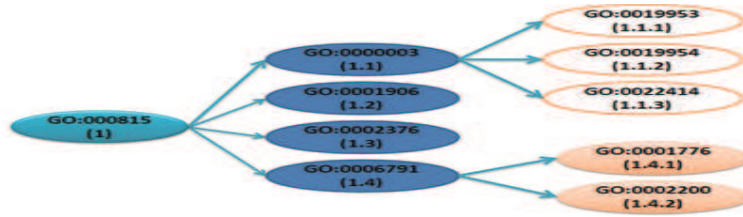


Figure 10: The Directed Acyclic Graph (DAG) derived from GO.

In GO database, a set of vocabulary terms are defined as gene functions, and organized with different relationships, such as "is-a" relationship between parent and child and "part-of" relationship between part and whole. All terms are organized into a Directed Acyclic Graph (DAG) with terms as nodes and relationships as edges as shown in Figure 10. To define the functional similarity between two GO terms, the depth of the common deepest ancestor of these two terms is used. For example, for biological process, "GO:000815" is the highest level, and its label is 1. Its child nodes are labeled as 1.1,1.2,1.3,..., respectively. "GO:0019953" and "GO:0019954" share the first two level and their GO similarity is 2. So the GO similarity of two drugs can be defined as follows.

$$GO(i, j) = \frac{\sum_{k,l} \max_g(GO_g(d_{ik}, d_{jl}))}{|C(d_i, d_j)|} \quad (3)$$

where $GO_g(d_{ik}, d_{jl})$ denotes the GO similarity between k th target of drug d_i and l th target of drug d_j with respect to a special process g , the maximum is used because one gene may be annotated by several GO terms, $|C(d_i, d_j)|$ is the size of all combinations of target proteins of drugs d_i and d_j . For a combination with n drugs, the GO similarity for drug components can be defined below.

$$G = \frac{\sum_{i,j} GO(i, j)}{C_n^2} \quad (4)$$

4 Conclusion

Drug combination is an attractive strategy for combating complex disease. In this work, a new network biology approach is presented to analyze the combination of drugs.

The target proteins of combination drugs are investigated in the context of protein interaction network, pathway and functional annotations. Furthermore, the molecular weights are investigated for drug components of combination drugs. By comparing positive combinations and negative combinations, it is found that the target proteins of positive combinations tend to be close in protein interaction network, and tend to be involved in functionally related pathways or biological processes. Furthermore, the positive combinations tend to share target proteins which also have more interaction partners compared with negative combinations. In addition, the drug components of the positive combinations tend to have similar molecular weights. These analysis results provides insights into mechanisms underlying drug combinations and can help to predict drug combinations in the future.

Acknowledgements

X.-M Zhao is partly supported by the Innovation Program of Shanghai Municipal Education Commission (10YZ01), Innovation Funding of Shanghai University and Shanghai Rising-Star Program. JS would like to thank the National Health and Medical Research Council of Australia (NHMRC) for financially supporting this research via the NHMRC Peter Doherty Fellowship.

References

- [1] Lindsay MA: Target discovery. *Nat Rev Drug Discov* 2003, 2: 831-838.
- [2] Yang K, Bai HJ, Ouyang Q, Lai LH, Tang C: Finding multiple target optimal intervention in disease-related molecular network. *Molecular Systems Biology* 2008, 4: 228.
- [3] Hopkins AL: Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008, 4: 682-690.
- [4] Geva-Zatorsky N, *et al.* Protein dynamics in drug combinations: a linear superposition of individual-drug responses. *Cell* 2010, 140(5): 643-651.
- [5] Calzolari D, *et al.* Search Algorithms as a Framework for the Optimization of Drug Combinations. *PLoS Comput Biol* 2008, 4(12): e1000249.
- [6] Rosenstock J, *et al.* Initial treatment with Rosiglitazone/Metformin fixed-dose combination therapy compared with monotherapy with either Rosiglitazone or Metformin in patients with uncontrolled type 2 diabetes. *Diabetes Obes Metab* 2006, 8: 650-660.
- [7] Agrawal N, *et al.* Identification of combinatorial drug combination regimens for treatment of Huntington's disease using *Drosophila*. *Proc Natl Acad Sci* 2005, 102(10): 3777-3781.
- [8] Wu Z, Zhao X.M, Chen L: A Systems biology approach to identify effective cocktail drugs. *BMC Systems Biology*, In Press.
- [9] Liu Y, *et al.* DCDB: Drug Combination Database . *Bioinformatics* 2010, 26(4): 587-588.
- [10] Wishart DS, *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008, 36: 901-906.
- [11] Stark C, *et al.* Biogrid: A General Repository for Interaction Datasets. *Nucleic Acids Res* 2006, 34: 535-539
- [12] Kanehisa M, *et al.* KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010, 38: 355-360.
- [13] Ashburner M, *et al.* Gene ontology: tool for the unification of biology. *Nat Genet* 2000, 25(1): 25-29.